Data Visualization (CIS 490/680)

Trees & Design

Dr. David Koop





What are the ingredients for a geospatial visualization?









Map Projection + Position Data

















Don't Just Create Population Maps!





D. Koop, CIS 680, Fall 2019

PET PEEVE #208: GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS









Cartograms



- Data: geographic geometry data & two quantitative attributes (one part-of-whole)
 - Derived data: new geometry derived from the part-of-whole attribute
- Tasks: trends, comparisons, part-of-whole
- How: area marks from derived geometry,
 - color hue/saturation/luminance
- Scalability: thousands of regions
- Design choices:
 - Colormap
 - Geometric deformation











Networks

- Why not graphs?
 - Bar graph
 - Graphing functions in mathematics
- Network: nodes and edges connecting the nodes
- Formally, G = (V,E) is a set of nodes V and a set of edges E where each edge connects two nodes.
- Nodes == items, edges connect items
- Both nodes and edges may have attributes









Node-Link Diagrams

- Data: nodes and edges
- Task: understand connectivity, paths, structure (topology)
- Encoding: nodes as point marks, connections as line marks
- Scalability: hundreds
- ... but high **density** of links can be problematic!
- We need position info—a layout!







Force-Directed Layout

- edges are springs that pull them together produce difference results each time it runs
- Nodes push away from each other but • Weakness: nondeterminism, algorithm may

D. Koop, CIS 680, Fall 2019









8

"Hairball"



JGD_Homology@cis-n4c6-b4. 26028 nodes, 100290 edges.









<u>Midterm</u>

- Thursday
- Covers material through this week
- Format:
 - Multiple Choice
 - Free Response (often multi-part)
 - CS 680 students will have extra que discussed

- CS 680 students will have extra questions related to the research papers





Project Design

- Start working on turning your visualization ideas into designs
- Sketch
- Options:
 - Try vastly different options
 - Refine an initial idea





Next Week

- Tuesday: Guest Lecture from Prof. Sun
- Thursday: No class, work on projects

D. Koop, CIS 680, Fall 2019

Sun sts





Hierarchical Edge Bundling











Hierarchical Edge Bundling



D. Koop, CIS 680, Fall 2019







14

Hierarchical Edge Bundling

- Flexible and generic method
- - information
 - explicit adjacency edges between their respective child nodes

 Reduces visual clutter when dealing with large numbers of adjacency edges Provides an intuitive and continuous way to control the strength of bundling. - Low bundling strength mainly provides low-level, node-to-node connectivity

- High bundling strength provides high-level information as well by implicit visualization of adjacency edges between parent nodes that are the result of









15

Bundling Strength













Adjacency Matrix

- Change network to tabular data and use a matrix representation
- Derived data: nodes are keys, edges are boolean values
- Task: lookup connections, find wellconnected clusters
- Scalability: millions of edges
- Can encode edge weight, too

D. Koop, CIS 680, Fall 2019











17

Cliques in Adjacency Matrices











Structures from Adjacency Matrices







Node-Link or Adjacency Matrix?

- Empirical study: For most tasks, node-link is better for small graphs and adjacency better for large graphs
- Multi-link paths are hard with adjacency matrices
- Immediate connectivity or neighbors are ok, estimating size (nodes & edges also ok)
- People tend to be more familiar with node-link diagrams
- Link density is a problem with node-link but not with adjacency matrices









Trees

- Trees are directed acyclic graphs
 - each edge has a direction: the origin is the parent, the destination is the child
 - cannot get back to a node after leaving it
- ...plus each node has at most one parent node
- A tree has a **root** (every other node hangs off it)
- Can consider enclosure in trees using parent-child relationships





21

Tree Visualizations



D. Koop, CIS 680, Fall 2019









Η

G







Node-Link Diagram

- Trees are graphs
- ...but we have more structure
- Horizontal or vertical
- Idea 1: partition space for each node via recursion
- Idea 2: "Tidy" Drawing
 - Wetherell & Shannon: Don't waste space (overlapping parent nodes is ok)
 - Reingold and Tilford: Keep symmetry, subtrees look similar



Reingold-Tilford Algorithm



- Recurse on left and right subtrees
- Shift subtree over as long as it doesn't overlap
- Place parent centered above the subtrees
- Originally, only binary trees, extended by Walker











Icicle Plot

- Line marks
- Vertical position shows depth
- Horizontal position shows links and sibling order
- Scalability: 1 pixel leaves, but harder to label













Radial Node-Link

- Use polar coordinates instead of rectilinear
- Same layout algorithms work (e.g. Reingold-Tilford)
- Benefit: space usage, labels









Sunburst

- Icicle plot in a radial layout
- Reading labels?
- Intuitive navigation







Indented Outline



D. Koop, CIS 680, Fall 2019

Like a filesystem tree

 Use horizontal position to show depth, vertical positions show sibling/order









Treemap

lode ink ree ayout 2,870		Circle Layout 9,317	Tree Map Layout 9,191	Stacl Area Layo 9,12 ⁻	ut La	prce rected yout 411	Data 20,544	L	oata ist 9,788		Strings 22,026		hapes 9,118	Mat 17,7		Displays 12,555	iff add no gte mod an ite or	59 59 59 1 stc xo va ic wi se di ave si c	Query 13,896		Max Flow Min Cut 7,840	Shortest Paths 5,914
adial ree ayout 2,348		Layout 7,881		lcicle Tree Layout	Bundl Edge Route	ec Indent Tree r Layou												d ma: 2 2 c min Match 3,748	Composi Expressi 3,677	t Expressic c Iterator 3,617	Link Distance 5,731 Betweenness	Spa Tree 3,41
Fircle acking ayout 2,003		Axis Layout 6,725		4,864 Dendrogra Layout 4,853	am Pie Layou	3,174 Ra It Lay	Node Sprite 19,382	Sc Bir 11,	ale I nding \$ 275 ·	Data Sprite 10,349	Color Palette 6,367	Size Shap Palett Pale 2,291 2,05	colors ti 10,001 9		Dates 8,217	Sort 6,887	Comparison 5,103 Date Util	Fn 3,240	Binary Express 2, 2,893	732 2,039	Centrality 3 534 Hierarchical Cluster 6,714	Com Struc 3,812
abeler ,956	3.899	Stack Area Label 3,202 Bifc	ortion 14 cal cortion	Fishe Disto 3,444	2,728 Operator List 4 5,248				Tree		Geometry 10,993	Palette	Sparse Matrix 3,366 Dense Matrix 3,165	I Matri 2,81		Filter 2,324	4,141 String Util 4,130 Arithmetic 3,891	Variance 1,876 Aggrega Express 1,616 Range	1,554 1 Literal 1,214 A Variable 1,124	And Or Dis ,02 970 933 Averag Sum Maximi Cou	Agglomerative Cluster 3,938 Aspect Ratio Banker	e Merg
roperty ncoder 138	Color Encoder 3,179	4,40	61 neye e er	Graph Distar Filter 3,165		2,023	Tree Builder 9,930 Edge Renderer	Shape	7,147		Fibonacci Heap 9,354	1,	e Arrays 8,258		Property 5,559	Orienta 1,486 I Value	Time Scale 5,833	1,594 Ordinal Scale 3,770	Log Scale 3,151	Ainimu 781 N Body Force 10,498		Partic Sj 2,822 2,
ncoder ,060 poltip ontrol 435	Shape Encoder Selection Control 7,862	Pan Zoom Contro 5,222	Legend 20,859		Operator 2,490	I Opera	5 569	Legend	Arr Tyj 69 Sprite J 3.301 Visualizati 16,540		Interpolator 8,746	Color Rec Interr Inter 2,047 2,04 Point I Interpola		ig O		Transition 9,201	Quantitative Scale 4,839 Scale 4,268	Quantile Scale 2,435	e Scale Roo Type Sca 1,82 [:] 1,75 Linear Scale	5 5 9,983	n	Sprinc G Force Fo 1,681 1, Drag Force
ver ntrol	Click Control 3,824	And Cor 2.13	ho tro Axis 38 24,593				Cartes Axes	ian			Matrix Interpolator Transitioner 19,975	Object I Interpola	1 320		Scheduler 5,593	Parallel 5,176	GraphML Converter 9,800	0 405	Scale ata Data burce Util 331 3,322	Text Sprite 10,066		Force Rect F Sprite V 3,623 4
ontrol st 665	Expand Control 2.832 Drag Control	2,10 Cor 1,38	tro 53				6,703 Axes 1,302	Axis Axis	Data Event Selection Event	Tooltip Event 1.701 Visualiz Event	2		Func Sequ 5,842	ence	Sequence 5,534	Transitio Event I Schedult	F 4 JSON I Converter Da 2 220 Co	ata 721 Fie	165 77			Line Sprite 1,732







Car/Truck Treemap

Truck Sales Slip, Tripping Up Chrysler

Over the past few years, Chrysler executives said they were following the lead of Toyota and Honda, focusing on vehicles that met the needs of their customers. But as American consumers turned away from large trucks and S.U.V.'s in 2006, Chrysler continued to churn out big vehicles, which are now sitting unsold at dealerships across the country.



General Motors

shifted to cars.

Cars

Dodge

Ram

Trucks/vans/S.U.V.'s 2.5 million 1.6 million Cars

G.M. introduced new versions of its large S.U.V.'s in late 2005, hoping they would bolster sales. Instead, sales of big vehicles were hurt when gas prices climbed. One of the few standouts was the Chevrolet HHR, new in 2005.



D. Koop, CIS 680, Fall 2019

Chevrolet

-8.7%



READING THE CHART

Boxes are scaled proportionally 25,000 according to number of cars 100,000 sold in 2006



Many of these vehicles were introduced in 2005.

lodge Saliber 12,224		
kge ktus 993		

Toyota Tacoma 178,351	Toyota Sienna 163,269	Toyota Camry 448,645		Toyota Corola 387,388	1	
Toyota RAV4 152,047	Toyota Highlander 129,794					
Toyota Tundra 124,508	Toyota 4Runner 103,086	Toyota Prius 106,971	Scion tC 79,125	Toyota Yari 70,308		on x8 306
Lexus FIX 108,348	Toyota FJ Cruiser 56,225 24,315	Toyota Avalon 88,938	Lexus ES 75,987	54,267	Scion xA 32,603	Lexus GS 27,390



Toyota Trucks/vans/S.U.V.'s 1.1 million Cars

Toyota rolled out a new version of the Camry, and once again it was the country's best-selling car. Toyota



Corolla sales also jumped, along with gas prices. Toyota could not escape the decline in sales of supersized S.U.V.'s like its Sequoia.

Honda

Trucks/vans/S.U.V.'s 0.7 million Cars 0.8 million

Like the Corolla, the small Honda Civic did well. But the Accord stalled. Buyers, it seems, are waiting for the new version to be released this year.



Northern Illinois University





+3.2%

Corolla

1.5 million

+12.5%

No 2005 sales

Car/Truck Treemap

Ford

-8.3%

Trucks/vans/S.U.V.'s 1.8 million Cars 1.1 million

Even the country's best-selling vehicles, the F-Series, slumped in 2006, with sales dropping 13 percent. One of Ford's bright spots was the new Fusion sedan, which made its debut in late 2005 and sold well in its first full year.

Ford F-Series

Ford F-Series 744,996	Ford Econolina 180,457			1 Expl ,229	anar	Ford Focus 177,006		Ford Taurus 174,803
	Ford Escape 157,395			Fore 92,4	i Ranger I20	Ford Mustang 166,530		Ford Fusion 142,502
	Ford Expedition 87,203	Ford F 50,125		itar	Morcury Mariner 33,941	Ford Five Hundred 84,218	Mercury Grand Marquis	Marcury Milan 35,653
		Volvo XC90 33,200	1	Lincoli Naviga 23,947	ator		54,685	Volvo S40 24,566
	Ford Freestyle 58,602	Mercu Mount 29,567	tainex	ө		Ford Crown Victoria 62,976	Lincoln Town Ca 39,295	-

READING THE CHART





D. Koop, CIS 680, Fall 2019



[A. Cox and H. Fairfield, <u>NYTimes</u>, 2012]











Treemap

- Containment marks instead of connection marks • Encodes some attribute of the items as the **size** of the rectangles Not as easy to see the intermediate rectangles

- Scalability: millions of leaf nodes and links possible
- Need a layout algorithm!







Layout Algorithms

- How do we generate the area marks?
- What considerations should we try to keep in mind?







Layout Algorithms

- How do we generate the area marks?
- What considerations should we try to keep in mind?
 - area true to quantitative value
 - show hierarchy
 - aspect ratio
- Also...
 - ordering
 - stability

D. Koop, CIS 680, Fall 2019

ks? to keep in mind?







Layout Algorithms

- Aspect ratio concerns: square is better
- <u>Slice and dice</u>:
 - Split at each level into strips
 - At each step, orientation of division (horiz/vert) changes
- Strip

- Order rectangles and move to a new row when aspect ratio gets worse







Improving Treemaps (Squarified)

- but can lead to bad aspect ratios
- rectangles
- to maintain good aspect ratios
 - use left and right side
 - process large rectangles first
- Ordering not preserved which may cause issues if the data is updated

D. Koop, CIS 680, Fall 2019

Switching from horizontal to vertical cuts may be ok for nicely-behaved trees,

• Problem: harder to compare sizes, more difficult to select/mouse over the

• Solution: Choose divisions (x/y) based on the width/height of region in order









Squarification Algorithm











Squarification Algorithm













Squarified Treemaps



(a) File system

D. Koop, CIS 680, Fall 2019

(b) Organization













Squarified Layout

- Sort values
- Switch orientation whenever necessary to obtain best aspect ratios





Improving Treemaps (Cushion)

- Leaves are ok, but it can be difficult to find the hierarchy
- Encode this as shading information
- More effective to understand hierarchy



D. Koop, CIS 680, Fall 2019



[van Wijk and van de Wetering, 1999]



Northern Illinois University





Disk Inventory

om In Zoom Out Mo	ve To Trash	Show Package Contents
Name	Size	
V 🗊 Contents	31,3 MB	
▶ 📁 Resources	21,5 MB	
NetServices	5,7 MB	
WacOS	2,2 MB	
iPhoto	1,9 MB	
iPhotoDPA	273 kB	
photocd	70 kB	
🕨 🧊 Plugins	1,7 MB	
.DS_Store	6 kB	
PlugIns Disabl	6 kB	
Info.plist	1 kB	
version.plist	463 Byte	
PkgInfo	8 Bytes	
Resources Dis	0 Bytes	
lcon	65 kB	
DS_Store	бКВ	
DS_Store hoto (iPhoto/Content	6 kB	

Color	Kind	Size	Files
-	Interface Builder Docum	15,4 MB	2104
-	MP3 Audio File	4,8 MB	2
	Unix Executable File	3,8 MB	23
	JPEG Image	1,6 MB	74
	Strings File	1,4 MB	348
-	HTML document	1,3 MB	333
-	TIFF Document	1,0 MB	310
-	Document	886 kB	16
-	Portable Network Graphi	635 kB	21
-	XML Property List File	183 kB	332
-	Apple Icon Image	109 kB	2
-	AIFF Audio	67 kB	2
	Finder Document	65 kB	1
	Script	35 kB	5
	Rich Text Format (RTF) d	30 kB	2
	AppleScript Suite Definit	7 kB	1
	AppleScript Suite Termin	6 kB	1
	Graphics Interchange Fo	5 kB	12
	Cascading Style Sheet (C	4 kB	4
	Symbolic Link	164 Byte	9









Squarified + Cushioned Treemaps



(a) File system

D. Koop, CIS 680, Fall 2019

(b) Organization







Nested Circles

- Looks more like cluster diagram, but shows hierarchy
- Containment shown by the layering of semi-transparent circles
- Labeling becomes more difficult

D. Koop, CIS 680, Fall 2019







44

Compound Networks

- Add a hierarchy to the network (e.g. from clustering)
- GrouseFlocks: uses nested circles with colors



D. Koop, CIS 680, Fall 2019

NIU



